arXiv:0903.2487v1 [hep-ph] 16 Mar 2009

# Are $BR(\overline{B} \to X_s\gamma)$ and $(g-2)_\mu$ consistent within the Constrained MSSM?

**Farhan Feroz**

*Astrophysics Group, Cavendish Laboratory, University of Cambridge*
*J.J. Thomson Avenue, Cambridge, CB3 0HE, UK*
*E-mail:* ff235@mrao.cam.ac.uk

**Mike Hobson**

*Astrophysics Group, Cavendish Laboratory, University of Cambridge*
*J.J. Thomson Avenue, Cambridge, CB3 0HE, UK*
*E-mail:* mph@mrao.cam.ac.uk

**Leszek Roszkowski**

*Department of Physics and Astronomy, University of Sheffield,*
*Sheffield S3 7RH, England E-mail:* L.Roszkowski@sheffield.ac.uk

**Roberto Ruiz de Austri**

*Instituto de Física Corpuscular, IFIC-UV/CSIC*
*Valencia, Spain*
*E-mail:* rruiz@ific.uv.es

**Roberto Trotta**

*Astrophysics Group, Imperial College London*
*Blackett Laboratory, Prince Consort Road, London SW7 2AZ, UK*
*E-mail:* r.trotta@imperial.ac.uk

ABSTRACT: We employ two different statistical tests to examine whether, in the framework of the Constrained MSSM, the experimentally determined values of $BR(\overline{B} \to X_s\gamma)$ and the anomalous magnetic moment of the muon, $(g-2)_\mu$ are consistent with each other. Our tests are designed to compare the theoretical predictions of the CMSSM in data space with the actual measurements, once all of the CMSSM free parameters have been integrated out and constrained using all other available data. We investigate the value of $(g-2)_\mu$ as obtained by using $e^+e^-$ data alone (which shows a $\sim 3\sigma$ discrepancy with the Standard Model prediction) and as obtained based on $\tau$ decay data (which shows a much milder, $\sim 1\sigma$ discrepancy). We find that one of our tests returns either a statistically inconclusive result or shows weak evidence of tension between $BR(\overline{B} \to X_s\gamma)$ and the $e^+e^-$–data based value of $(g-2)_\mu$. On the other hand, our second test, which is more stringent in this application, reveals that the joint observations of $BR(\overline{B} \to X_s\gamma)$ and $(g-2)_\mu$ from $e^+e^-$ data alone are incompatible within the CMSSM at the $\sim 2\sigma$ level. On the other hand, for both tests we find no significant tension between $BR(\overline{B} \to X_s\gamma)$ and the value of $(g-2)_\mu$ evaluated using $\tau$ decay data. These results are only weakly dependent on the three different priors that we employ in the analysis. We conclude that, if the discrepancy between the Standard Model and the experimental determinations of $(g-2)_\mu$ is confirmed at the $\sim 3\sigma$ level, this could be interpreted as strong evidence against the CMSSM.

KEYWORDS: Supersymmetric Effective Theories, CMSSM, anomalous magnetic moment, statistical tests.

## Contents

## 1. Introduction

Softly broken low-energy supersymmetry (SUSY) is considered to be perhaps the most promising theory beyond the Standard Model (SM). Not only does it provides an elegant solution to the hierarchy problem [1] but also naturally accommodates gauge coupling unification [2] and offers a clue to the dark matter (DM) problem in the Universe [3].

On the other hand, without specifying a complete underlying mechanism of SUSY breaking, the general Minimal Supersymmetric Standard Model (MSSM) suffers from a large number of SUSY-breaking soft parameters which are poorly determined. Motivated by a natural link between SUSY and grand unified theories (GUTs), over the last several years it has become customary to impose various boundary conditions at the GUT scale and explore resulting SUSY phenomenology. The most popular model of this class is the Constrained MSSM (CMSSM) [4], includes the minimal supergravity model (mSUGRA) [5]. In this scheme one defines all SUSY parameters at the unification scale $M_{\mathrm{GUT}}$ and next employs the Renormalization Group Equations (RGEs) to evolve them down and compute the

couplings and masses in an effective theory valid at the electroweak scale. The CMSSM is defined in terms of four continuous free parameters: common scalar ($m_0$), gaugino ($m_{1/2}$) and tri-linear ($A_0$) mass parameters (all specified at the GUT scale), plus the ratio of Higgs vacuum expectation values $\tan\beta$; and one discrete parameter $\mathrm{sgn}(\mu)$, where $\mu$ is the Higgs/higgsino mass parameter whose square is computed from the conditions of radiative electroweak symmetry breaking (EWSB).

The phenomenology of the CMSSM has been studied in a vast number of papers. The usual approach has been to explore the model by performing fixed grid scans in $m_{1/2}$ and $m_0$ for fixed, "representative" values of $\tan\beta$ and for $A_0 = 0$ [6], and also for fixed values of SM parameters, e.g., the top mass $m_t$, which can have a large impact on results especially at large $m_0$. Also, the model's predictions for observable quantities, e.g., the relic abundance $\Omega_\chi h^2$ of the neutralino, or Higgs and superpartner masses have been compared with experimental data in a simplified way: if the predicted values is within some arbitrary range, typically $1\sigma$ or 90% CL then the point is treated as "allowed"; otherwise it is rejected. Theoretical errors are also typically ignored. A $\chi^2$ approach applied in [7, 8, 9] addressed the latter problems. On the other hand, in those papers it was advocated to reduce the effective number of CMSSM parameters by using the well-measured value of $\Omega_\chi h^2$ to determine a "surface" in the model's parameter space was somewhat questionable as its shape and "thickness" can critically depend on the actual value of $m_t$, especially at large $m_0$ (compare fig. 4 in [10]) and also by including a "fudge factor" in the definition of $\chi^2$ in order to suppress the contribution of the large $m_0$ region (see eq. (1) in [8]). In a more recent analysis [11] that element of the $\chi^2$ analysis has been abandoned. On the other hand, the conclusions of [11] heavily rely on the somewhat uncertain discrepancy between the SM and the experimental determinations of the anomalous magnetic moment of the muon $(g-2)_\mu$.

Over the last few years a new approach based on Bayesian statistics linked with either Markov Monte Carlo Chain (MCMC) [12, 13, 14, 15, 16, 17, 10, 18, 19, 20, 21, 22] or Nested Sampling (NS) scanning methods [23, 24, 25] has been successfully applied in a well defined statistical framework (see, e.g., [26]). Furthermore the priors issue, considered as a "soft spot" of the Bayesian approach, has been recently thoroughly addressed and has been shown to embody in a quantitative manner the physical fine-tuning of the theory, see [27].

One of the outcomes of the most recent and more sophisticated scans has been to realize that even the CMSSM, with its relative economy of free parameters, remains presently somewhat underconstrained by currently available data, thus leaving large regions of CMSSM parameters allowed [25]. Despite this, it was pointed out in [10], and investigated in more detail in [25], that there appears to exist a certain "tension" between the current measurements of $BR(\overline{B} \to X_s\gamma)$ (hereafter denoted by $b \to s\gamma$ for brevity) and $(g-2)_\mu$, in the sense that the two observables favor different regions of the CMSSM parameters space (compare figs. 8 and 10 in [25]). This is because the $BR(\overline{B} \to X_s\gamma)$ constraint favors the focus point (FP) region [28, 29], as the (always positive) charged Higgs/top contribution has to be large enough so that, starting from the SM central value of $3.12 \times 10^{-4}$, the (negative, for $\mu > 0$) chargino/stop contribution can bring the sum down to the experimental central value of $3.55 \times 10^{-4}$. This requires the charged Higgs to be

light enough and the stop (or chargino, or both) to be heavy enough. Both conditions are satisfied in the FP region. On the other hand, large corrections to the $(g-2)_\mu$ values arise mostly in the low–mass region. We feel that it would be interesting to further investigate this tension, and to develop statistical tools to quantify the possible incompatibility of the two observations within the theoretical model. A strong tension between the data would then be interpreted either as a sign of an undetected (or underestimated) systematic error in one of the data sets, or as a sign that the theoretical model is at odds with the data and hence is disfavored. One first evaluation of the tension has been carried out in [24] using a model comparison test, returning however an inconclusive result. The purpose of this paper is to re-consider the problem of the tension between these two observables by addressing it with a novel statistical test, called "the predictive likelihood ratio test", or the $\mathscr{L}$–test introduced below. We clarify that the reason why the test based on model comparison performed in [24] is inconclusive can be traced back to the orientation of degeneracies in data space, a feature that in this particular context makes the model comparison test less stringent than the new test introduced here.

The paper is organized as follows. In section 2 we introduce the statistical framework and define the new test for the compatibility of observables based on the predictive data distribution (the $\mathscr{L}$–test) as well as present the test based on the model comparison (the $\mathscr{R}$–test). In sec. 3 we specify the theoretical model and its parameters and the priors we consider and we apply the statistical tests to the CMSSM. We present our numerical results in section 4. Our conclusions are given in section 5.


## 2. Setup

### 2.1 Statistical framework

We follow here the notation and conventions of our previous works [15, 17, 25]. We denote the set of parameters of the model $\mathcal{M}$ under consideration (here the CMSSM) by $\theta$, and by $\psi$ all other relevant parameters, the so-called *nuisance parameters*, which here include relevant SM quantities. Both sets form our *basis parameters*

$$m = (\theta, \psi). \tag{2.1}$$

Bayesian inference is based on Bayes' theorem which reads

$$p(m|d, \mathcal{M}) = \frac{p(d|m, \mathcal{M})p(m|\mathcal{M})}{p(d|\mathcal{M})}. \tag{2.2}$$

The quantity $p(m|d, \mathcal{M})$ on the l.h.s. of eq. (2.2) is called a *posterior probability density function* (posterior pdf, or simply a *posterior*). On the r.h.s., the quantity $p(d|m, \mathcal{M})$, taken as a function of $m$ for *fixed data $d$*, is called the *likelihood*. The likelihood supplies the information provided by the data. The quantity $p(m|\mathcal{M})$ denotes a *prior probability density function* (prior pdf, or simply a *prior*) which encodes our state of knowledge about the values of the parameters in $m$ before we see the data. The prior state of knowledge is then updated to the posterior via the likelihood.

Finally, the quantity in the denominator is called *the evidence* or *model likelihood*, which is obtained by computing the average of the likelihood under the prior (so that the r.h.s. of eq. (2.2) is properly normalized to unity probability),

$$p(d|\mathcal{M}) = \int p(d|m, \mathcal{M})p(m|\mathcal{M})\mathrm{d}m. \tag{2.3}$$

If one is interested in constraining the model's parameters, the evidence is merely a normalization constant, independent of $m$, and can therefore be dropped. However, the evidence is very useful in the context of Bayesian model comparison (see e.g. [30, 31] and [16, 24] for recent applications to the CMSSM). One of the main goals of this paper is to develop and apply to the CMSSM a new evidence-based statistical test on the consistency of two or more observables within a given theoretical model. Taking into account that one might wish to consider different models, all of the above relations have been conditioned explicitly on the model under consideration, $\mathcal{M}$. However, in the following we will drop the explicit conditioning on $\mathcal{M}$ since we only work in the framework of a single, given model in this paper, namely the CMSSM.

## 2.2 The predictive likelihood ratio test ($\mathscr{L}$–test)

Let us split the full data set of $n$ observables $d$ (which will be given below) as $d = \{\mathscr{D}, D\}$. Suppose and that we are interested in testing the compatibility of the observations within a subset $\mathscr{D} = \{\mathscr{D}_1, \ldots, \mathscr{D}_k\}$, $k < n$, conditional on the observed values for the second part of the data set, $D = \{d_{k+1}, \ldots, d_n\}$, which are considered as external (independent) constraints which are assumed to be correct. We are thus interested in evaluating the *conditional evidence $p(\mathscr{D}|D)$*, which represents the probability of measuring data $\mathscr{D}$ given that data $D$ have been gathered for the remaining $n - k$ observables. In other words, this conditional probability can be interpreted as the predictive probability for a measurement of the observables $\mathscr{D}$ given what has been observed for the other quantities. As a consequence of the basic probability manipulation rules, the conditional evidence can be written as

$$p(\mathscr{D}|D) = \frac{p(\mathscr{D}, D)}{p(D)}, \tag{2.4}$$

(recall that we are dropping the conditioning on the model $\mathcal{M}$ which is understood). On the r.h.s., the joint evidence $p(\mathscr{D}, D)$ is the probability of measuring the joint data set within the assumed model, *independently of the actual true values of the model's parameters $m$*, which have been integrated out in the computation of the evidence, see eq. (2.3). The joint evidence has to be evaluated *as a function of the possible outcomes of the observations of the data set $\mathscr{D}$*. This requires evaluating the evidence for a series of possible values for $\mathscr{D}$, at each time integrating over the full parameter space of the model. The possible data realizations $\mathscr{D}$ are different outcomes for the measurements (e.g., different means) given the experimental noise, i.e., the reported error of the central value. At the same time, the data set $D$ are held fixed at their actual observed values, for, as stated above, we assume that this part of the data set is trustworthy and can be used to constrain the model's parameters. Notice that while the central values of the data set $D$ are assumed

to be correct, the uncertainty on their value is automatically fully accounted for, since we integrate over all the model's parameters when computing the evidence and we include both experimental and theoretical errors on $D$.

Once $p(\mathscr{D}|D)$ is obtained as a function of $\mathscr{D}$, its evaluation at the observed value $\mathscr{D} = \mathscr{D}^{\mathrm{obs}}$ allows one to determine the compatibility of the observed data realization $\mathscr{D}^{\mathrm{obs}}$ with the model and the rest of the data, by evaluating the relative probability of obtaining such a realization compared to the maximum probability for the data set in question. Let us denote by $\mathscr{D}^{\mathrm{max}}$ the values of the data that maximises $p(\mathscr{D}|D)$. Then the relevant quantity to consider is the ratio

$$\mathscr{L}(\mathscr{D}^{\mathrm{obs}}|D) \equiv \frac{p(\mathscr{D}^{\mathrm{obs}}|D)}{p(\mathscr{D}^{\mathrm{max}}|D)} = \frac{p(\mathscr{D}^{\mathrm{obs}}, D)}{p(\mathscr{D}^{\mathrm{max}}, D)}, \qquad (\mathscr{L}\text{–test}), \qquad (2.5)$$

where we have used eq. (2.4) in the second equality. This is analogous to a likelihood ratio in data space, *but integrated over all possible values of the parameters of the model.* We call the $\mathscr{L}$–test the *predictive likelihood ratio test.* If $\mathscr{L}(\mathscr{D}^{\mathrm{obs}}|D) \sim 1$, this shows that both data sets are compatible with each other and with the model's assumptions (including the prior choice), and therefore we can legitimately use them together to constrain the parameters of the model $\mathcal{M}$. If however $\mathscr{L}(\mathscr{D}^{\mathrm{obs}}|D) \ll 1$, we should doubt the consistency of the data $\mathscr{D}$ (perhaps considering the possibility of systematic effects) or the model's assumptions (i.e., the choice of model or of the assumed form and/or ranges of its priors). If $\mathscr{L}$–test comes out to be weakly dependent on the prior, then this will give us more confidence that the conclusions of the statistical test when applied to the assumed model are robust.

A simple example of the application of the $\mathscr{L}$–test method to a toy linear model is presented in Appendix A.

## 2.3 The model comparison test ($\mathscr{R}$–test)

A different compatibility test has been employed by [24], following earlier applications in cosmology [32]. The gist of what was called "model comparison test" there can be summarized as follows (see [24] for full details).

The idea is to perform a Bayesian model comparison test between two hypotheses, namely $\mathcal{H}_0$, stating that the data $\mathscr{D}$ under scrutiny are all compatible with each other and with the model, versus $\mathcal{H}_1$, purporting that the observables are incompatible (within the assumed model) and hence tend to pull the constraints in different regions of parameter space. For $k > 1$, the Bayes factor between the two hypotheses, giving the relative probabilities (odds) between $\mathcal{H}_0$ and $\mathcal{H}_1$ is given by

$$\mathscr{R} = \frac{p(\mathscr{D}|D, \mathcal{H}_0)}{\prod_{i=1}^{k} p(\mathscr{D}_i|D, \mathcal{H}_1)}. \qquad (2.6)$$

Writing again the conditional evidences in terms of the joint evidences, e.g. $p(\mathscr{D}|D, \mathcal{H}_0) = p(\mathscr{D}, D|\mathcal{H}_0)/p(D|\mathcal{H}_0)$, and noting that $p(D|\mathcal{H}_0) = p(D|\mathcal{H}_1)$ (which follows because the evidence from the data we are not testing does not depend on the hypothesis being considered), eq. (2.6) can be recast as

$$\mathscr{R}(\mathscr{D}) = \frac{p(\mathscr{D}, D|\mathcal{H}_0)}{\prod_{i=1}^{k} p(\mathscr{D}_i, D|\mathcal{H}_1)} p(D|\mathcal{H}_0)^{k-1} \qquad (\mathscr{R}\text{–test}, k > 1). \qquad (2.7)$$

If instead $k = 1$, i.e., $\mathscr{D} = \mathscr{D}_1$ and we wish to test the consistency of one single new observation, then eq. (2.7) needs to be modified to

$$\mathscr{R}(\mathscr{D}_1) = \frac{p(\mathscr{D}_1, D|\mathcal{H}_0)}{p(\mathscr{D}_1|\mathcal{H}_1)p(D|\mathcal{H}_1)} \qquad (\mathscr{R}\text{--test}, k = 1). \qquad (2.8)$$

Eqs. (2.7) and (2.8) are then evaluated at the observed value of the data sets being tested, i.e. for $\mathscr{D} = \mathscr{D}^{\text{obs}}$. If $\ln \mathscr{R}(\mathscr{D}^{\text{obs}}) > 0$, this is evidence in favour of the hypothesis $\mathcal{H}_0$ that the data are compatible. If instead $\ln \mathscr{R}(\mathscr{D}^{\text{obs}}) < 0$ the alternative hypothesis $\mathcal{H}_1$ that there is a tension among the data (and the model) is preferred. More quantitatively, the strength of evidence for either case can be assessed against so-called "Jeffreys' scale", which we report in table 1 along with our (slightly modified) convention for denoting the different levels of evidence.

| $|\ln \mathscr{R}|$ | Odds | Strength of evidence |
|---|---|---|
| < 1.0 | $\lesssim 3 : 1$ | Inconclusive |
| 1.0 | $\sim 3 : 1$ | Weak evidence |
| 2.5 | $\sim 12 : 1$ | Moderate evidence |
| 5.0 | $\sim 150 : 1$ | Strong evidence |

**Table 1:** Empirical scale for evaluating the strength of evidence (so-called "Jeffreys' scale"). Threshold values are empirically set, and they occur for values of the logarithm of the Bayes factor between the hypotheses of $|\ln \mathscr{R}| = 1.0$, 2.5 and 5.0. The right-most column gives our convention for denoting the different levels of evidence above these thresholds, according to the prescription in [33].

In applying the test to the CMSSM below we will consider the cases $k = 1$ and $k = 2$ with, as mentioned above, the two pieces of data being tested for mutual consistency being $b \to s\gamma$ and $\delta(g-2)_\mu$ (the latter both from $\tau$ decay and $e^+e^-$ data separetely).

## 3. An application to the CMSSM

Before we apply the above formalism to the CMSSM, we first specify the priors tested and the experimental constraints.

### 3.1 Choice of priors and data

In order to assess the robustness of our results with respect to plausible changes of priors, we consider three different classes of priors:

- **flat prior:** flat on $m_0, m_{1/2}, A_0, \tan\beta$, with ranges as given in section 3.2 of [25];

- **log prior:** flat on $\ln m_0, \ln m_{1/2}, A_0, \tan\beta$, with ranges as given in section 3.2 of [25];

- **CCR mSUGRA prior:** flat on $m_0, m_{1/2}, A_0, B$ but with an effective "penalty term" that naturally leads to low fine tuning among SUSY parameters.

Unlike the first two priors, which refer to the CMSSM parameterization in terms of its parameters $\theta = (m_{1/2}, m_0, A_0, \tan\beta)$, the third prior, as introduced in ref. [27] by Cabrera, Casas and Ruiz de Austri (hence the name), is applied to mSUGRA with its basic parameters $m_{1/2}, m_0, A_0, B$, augmented by the top Yukawa coupling $y_t$. A marginalization over $\mu$ selects the value $\mu_0$ that reproduces the experimental value of $M_Z$. As shown in ref. [27], it is also convenient and natural to trade the parameter $B$ for $\tan\beta$ and $y_t$ for the top mass $m_t \propto y_t \sin\beta$. This procedure results in an effective prior

$$p_{\text{eff}}(m_t, m_0, m_{1/2}, A_0, \tan\beta) = J|_{\mu=\mu_0} p(y_t, m_0, m_{1/2}, A_0, B, \mu = \mu_0). \qquad (3.1)$$

Assuming a flat prior on the parameters $m_{1/2}, m_0, A_0, B$ and a log prior on $y_t$, the Jacobian term acts as an effective "penalty term" that favors lower values of $\mu$ and $\tan\beta$, and thus leads to less fine-tuning as in the focus point region. In the CMSSM this corresponds to large $m_0$. On the other hand, the changing of parameters from $B$ to $\tan\beta$ favors large $m_{1/2}$ because of the $B$ dependence on $m_{1/2}$ in the RGEs. (See ref. [27] for more details.)

As we shall see, our results are largely insensitive to the choice of priors, which indicates a remarkable robustness of this statistical test. This can be traced backed to the fact that the parameters within the model are fully integrated out in the computation of the predictive probability.

| Observable | Mean value $\mu$ | Uncertainties $\sigma$ (exper.) | $\tau$ (theor.) | ref. |
|---|---|---|---|---|
| $\delta a_\mu^{\text{SUSY}} \times 10^{10}$ | 29.5 | 8.8 | 1.0 | [34] ($e^+e^-$ data) |
|  | 8.9 | 9.5 | 1.0 | [35] ($\tau$ data) |
| $BR(\overline{B} \to X_s\gamma) \times 10^4$ | 3.55 | 0.26 | 0.21 | [36] |

**Table 2:** Summary of the observables $\mathcal{D}$ being tested for consistency.

The focus of this paper is to test for consistency the measured values of $b \to s\gamma$ and the anomalous magnetic moment of the muon, $(g-2)_\mu$. For the latter, we consider two sets of measurements: the first is based on $e^+e^-$ data, and it gives a $\sim 3.2\sigma$ discrepancy with the SM predicted value [34]; the second one employs $\tau$ decay data to evaluate the SM hadronic contribution to $(g-2)_\mu$ instead, which leads to a much better agreement, $\delta a_\mu^{\text{SUSY}} = (8.9 \pm 9.5) \times 10^{-10}$ [35]. These values and their uncertainties are listed in the top part of Table 2.

As regards $BR(\overline{B} \to X_s\gamma)$, for the new SM prediction we obtain the value of $(3.12 \pm 0.21) \times 10^{-4}$.[1] We compute SUSY contribution to $BR(\overline{B} \to X_s\gamma)$ following the procedure outlined in refs. [39, 40] which was extended in refs. [41, 42] to the case of general flavor mixing. In addition to full leading order corrections, we include large $\tan\beta$-enhanced terms arising from corrections coming from beyond the leading order and further include (subdominant) electroweak corrections.

---

[1]The value of $(3.15 \pm 0.23) \times 10^{-4}$ originally derived in ref. [37, 38] was obtained for slightly different values of $M_t$ and $\alpha_s(M_Z)^{\overline{MS}}$. Note that, in treating the error bar we have explicitly taken into account the dependence on $M_t$ and $\alpha_s(M_Z)^{\overline{MS}}$, which in our approach are treated parametrically. This has led to a slight reduction of its value.

| Observable | Mean value | Uncertainties | | ref. |
| | $\mu$ | $\sigma$ (exper.) | $\tau$ (theor.) | |
| Nuisance paramaters | | | | |
| $M_t$ | 172.6 GeV | 1.4 GeV | N/A | [47] |
| $m_b(m_b)^{\overline{MS}}$ | 4.20 GeV | 0.07 GeV | N/A | [48] |
| $\alpha_s(M_Z)^{\overline{MS}}$ | 0.1176 | 0.002 | N/A | [48] |
| $1/\alpha_{\rm em}(M_Z)^{\overline{MS}}$ | 127.955 | 0.03 | N/A | [49] |
| Observables (measured) | | | | |
| $M_W$ | 80.398 GeV | 25 MeV | 15 MeV | [50] |
| $\sin^2\theta_{\rm eff}$ | 0.23153 | $16 \times 10^{-5}$ | $15 \times 10^{-5}$ | [50] |
| $\Delta M_{B_s}$ | 17.77 ps$^{-1}$ | 0.12 ps$^{-1}$ | 2.4 ps$^{-1}$ | [43] |
| $BR(\overline{B}_u \to \tau\nu) \times 10^4$ | 1.32 | 0.49 | 0.38 | [36] |
| $\Omega_\chi h^2$ | 0.1099 | 0.0062 | $0.1\,\Omega_\chi h^2$ | [44] |
| Observables (limits) | | | | |
| | Limit (95% CL) | | $\tau$ (theor.) | ref. |
| $BR(\overline{\overline{B}}_s \to \mu^+\mu^-)$ | $< 5.8 \times 10^{-8}$ | | 14% | [45] |
| $m_h$ | $> 114.4$ GeV (SM-like Higgs) | | 3 GeV | [46] |
| $\zeta_h^2$ | $f(m_h)$ (see ref. [15]) | | negligible | [46] |
| $m_{\tilde{q}}$ | $> 375$ GeV | | 5% | [48] |
| $m_{\tilde{g}}$ | $> 289$ GeV | | 5% | [48] |
| other sparticle masses | As in table 4 of ref. [15]. | | | |

**Table 3:** Summary of the observables $D$ used in the analysis, on which the consistency test is conditional. Upper part: measurements on nuisance (SM) parameters. (N/A stands for "not applicable".) Central part: Observables for which a positive measurement has been made. Lower part: Observables for which only limits currently exist. For details, see the treatment in ref. [15, 17, 25].

All the other experimental values of the collider and cosmological observables that we assume in order to perform the compatibility test for $\delta a_\mu^{\rm SUSY}$ and $BR(\overline{B} \to X_s\gamma)$ are listed in table 3. We refer to [15, 17, 25] for details about the computation of each quantity and for justification of the theoretical errors adopted, as well for a detailed description of the likelihood function. In particular, points that do not fulfil the conditions of radiative EWSB and/or give non-physical (tachyonic) solutions are discarded. Also, we take $\mu > 0$, because of its correlation with sign of $\delta(g-2)_\mu$.

### 3.2 Applying the $\mathscr{L}$–test to the CMSSM

We are interested in assessing the compatibility of $\mathscr{D} = \{b \to s\gamma, \delta(g-2)_\mu\}$, while assuming all the other data (denoted by $D$) to be believable. We remind the reader at this point that we are concerned with making predictions in data space, and *not* in parameter space, as it is usually done. We are not interested in constraining the parameters of the model here, but instead integrate over all their possible values. Therefore, the resulting values of $b \to s\gamma$ and $\delta(g-2)_\mu$ should be understood to represent the mean values that are predicted

to be obtained experimentally for the respective quantities within the CMSSM, once all the other constraints on the r.h.s. of the conditioning bar ($D$, as in table 3) are taken into account (including their experimental and theoretical uncertainties).

We thus evaluate the evidence and compute the predictive probability on a grid of values for $(b \to s\gamma, \delta(g-2)_\mu)$ representing the possible outcomes for the central value of the observation. At each point we keep the same experimental error as the one that has been effectively reported by the experiments (adding the theoretical error on top), as given in table 2. In other words, we consider different possible outcomes for the central values but with fixed instrumental noise properties, which is a reasonable assumption.

The computation of the evidence is numerically costly, as it involves an 8–dimensional integral over the whole parameter space for every choice of data values one wishes to test. We employ a modified version of the `SuperBayeS` package [17] including the MultiNest algorithm [25], which allows one to compute the evidence and from there, the predictive probabilities involved in the $\mathscr{L}$–test. Despite MultiNest's high efficiency, each evidence evaluation still requires about 3 days of CPU time on 4 3.00 GHz Intel Woodcrest processors. Appendix of ref. [25] provides a full description of how the uncertainty on the value of the evidence is evaluated with MultiNest. This uncertainty is then propagated to the uncertainty on the $\mathscr{L}$–test of eq. (2.5).

We scan over the following central values for the experimental outcomes, chosen to bracket the actually observed values:

$$BR(\overline{B} \to X_s\gamma) \times 10^4 : \quad 1.5, \ldots, 4.0 \ \text{ in intervals of } 0.5 \tag{3.2}$$

$$\delta(g-2)_\mu \times 10^{10} : \quad 0, \ldots, 40 \ \text{ in intervals of } 5. \tag{3.3}$$

As for the experimental noise, we fix this to the actually reported value for the real observation, supplemented by a suitable theoretical error, as given in table 2. When considering the two different experimental determinations of $\delta(g-2)_\mu$ (one based on $\tau$ decay data and one based on $e^+e^-$ data alone), we should in principle repeat our test using the reported experimental error for each of the observations. However, the reported experimental errors on $\delta(g-2)_\mu$ for the two determinations of the quantity are very similar (within about 10%) and therefore we employ the uncertainty reported in using the $e^+e^-$ data for both. This approximation is not expected to influence significantly our result.

## 4. Numerical results

It is interesting to consider both the $\mathscr{L}$–test and the $\mathscr{R}$–test, for each of them is sensitive to possible tensions between the observables in a different way and may in general give different results. (This is demonstrated in a toy model example in Appendix A.) This is in fact not surprising, for while different measurements can be compatible with each other and also compatible with the model being fitted in only one way if all the measurements are correct and the theory is the right model, there are many different ways in which an incompatibility could manifest itself. The $\mathscr{L}$–test asks what is the probability of measuring a certain value for the data subset $\mathscr{D}$ (relative to the maximum probability achievable under

the model) given what is known about the model from the remaining data $D$. The $\mathscr{R}$–test instead tries to enforce consistency between the data being tested and the remaining data sets, by looking for values of $\mathscr{D}$ that are jointly compatible with the parameter space singled out by $D$. These two approaches show subtle differencies and in general play out differently whenever a genuine tension between the observables exists.

Furthermore, the two tests are evaluated on different scales: the $\mathscr{L}$–test being of the form of a likelihood ratio test can be evaluated on a significance scale analogous to the usual $\Delta\chi^2$ rule, while the $\mathscr{R}$–test (representing odds between two hypotheses) should be assessed against Jeffreys' scale for the strength of evidence. Another issue to consider is that in general the two tests favour different degenerate regions of data space (see fig. 5 in Appendix A for an illustration) and one can easily imagine situations where one of these regions is more constrained than the other, due to the structure of the model. In this case, the test that exhibits more power along this more constrained degenerate region will appear to be more stringent.

## 4.1 Results for the $\mathscr{L}$–test

We begin by employing the $\mathscr{L}$–test to separately test the consistency between $\mathscr{D}_1 = b \to s\gamma$ and the other observables $D$ (but excluding from the latter $\delta(g-2)_\mu$) and between $\mathscr{D}_1 = \delta(g-2)_\mu$ and the other observables (but excluding from the latter $b \to s\gamma$). Notice in particular that we do include the dark matter constraint in the assumed data $D$. The outcome of these two tests is shown in fig. 1 and reported in table 4. In the left panel of fig. 1, we plot the quantity $\mathscr{L}(BR(\overline{B} \to X_s\gamma)|D)$ as a function of the possible outcome of the experimental observation, with the actual observed central value indicated by a vertical, solid line. In the right panel of fig. 1, we plot instead $\mathscr{L}(\delta(g-2)_\mu|D)$ as a function of the possible measured values of $\delta(g-2)_\mu$, indicating by vertical lines the actual observed values from $e^+e^-$ and $\tau$ data.

The CMSSM, once all the observations other than $\delta(g-2)_\mu$ are accounted for, tends to predict a $b \to s\gamma$ value close to the SM prediction, $BR(\overline{B} \to X_s\gamma) \simeq 3.12 \times 10^4$ (the precise value depending on the actual values of SM input parameters, especially $m_t$ and $\alpha_s(M_Z)^{\overline{MS}}$), with predominantly small negative corrections arising from chargino-stop loop contributions. This is shown by the peak in the predictive distribution, which occurs at around $BR(\overline{B} \to X_s\gamma) \sim 3 \times 10^4$. The experimental central value $3.55 \times 10^4$ is within about $1\sigma$ of the most likely value, thus it is not significantly in tension with the other observables (see top part of table 4).

Turning next to the anomalous magnetic moment of the muon (right panel of fig. 1), the predictive probability is largest for $\delta(g-2)_\mu \sim 0$, as might be expected from noting that only a small fraction of the CMSSM parameter space gives rise to sizable SUSY corrections to $\delta(g-2)_\mu$. The probability remains almost flat out to $\delta(g-2)_\mu \lesssim 10 \times 10^{-10}$, which means that the $\tau$ decay data determination is perfectly compatible with all other observations. Indeed, the $\mathscr{D}$ results in the bottom part of table 4 show that the results are not significant for the $\tau$ decay data. However, the predictive probability drops fairly steeply above that value (compare fig. 1), thus leading to tension for the $e^+e^-$ data, at about the $2\sigma$ level for
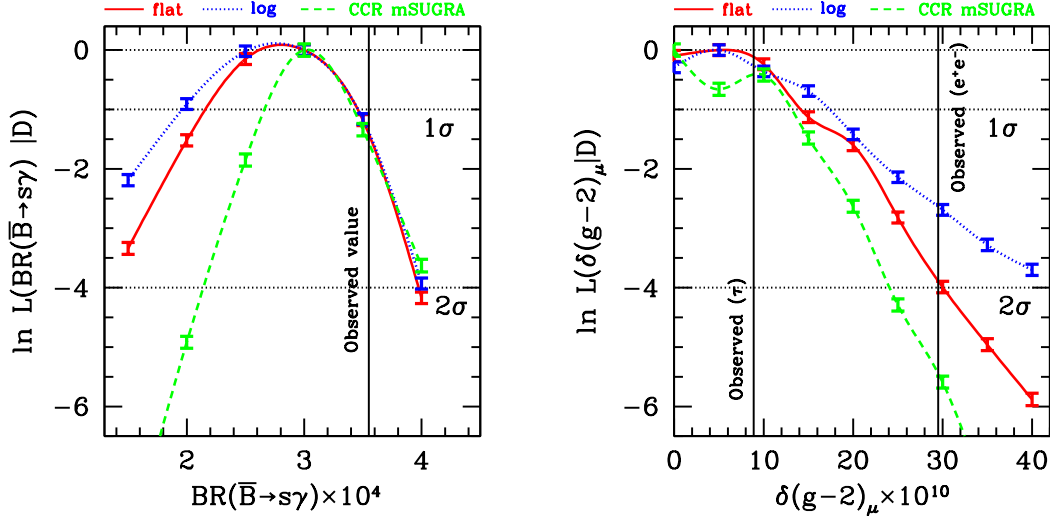
**Figure 1:** Predictive data distribution ($\mathscr{L}$–test) for $b \to s\gamma$ (left panel) and $\delta(g-2)_\mu$ (right panel) in the CMSSM for three different choices of priors: flat prior (red/solid), log prior (blue/dotted) and the CCR mSUGRA prior (green/dashed). The predictive distributions are conditional on all other observations, excluding $\delta(g-2)_\mu$ and $b \to s\gamma$. The vertical lines give the actual measured values. The errorbars denote the location at which the predictive probability has been computed (and its error), while the lines are a smoothed spline.

| Prior | $\ln \mathscr{L}(\mathscr{D}_1 \vert D)$ | Interpretation |
|---|---|---|
| $\mathscr{D}_1 = BR(\overline{B} \to X_s\gamma)$ | | |
| Flat | $-1.63 \pm 0.11$ | Not significant ($1.28\sigma$) |
| Log | $-1.43 \pm 0.12$ | Not significant ($1.20\sigma$) |
| CCR mSUGRA | $-0.89 \pm 0.13$ | Not significant ($< 1\sigma$) |
| $\mathscr{D}_1 = \delta(g-2)_\mu$ from $e^+e^-$ data | | |
| Flat | $-3.99 \pm 0.10$ | Incompatible at 95.4% significance |
| Log | $-2.69 \pm 0.10$ | Not significant ($1.64\sigma$) |
| CCR mSUGRA | $-5.59 \pm 0.10$ | Incompatible at 98.2% significance |
| $\mathscr{D}_1 = \delta(g-2)_\mu$ from $\tau$ decay data | | |
| Flat | $-0.24 \pm 0.10$ | Not significant ($< 1\sigma$) |
| Log | $-0.38 \pm 0.08$ | Not significant ($< 1\sigma$) |
| CCR mSUGRA | $-0.30 \pm 0.08$ | Not significant ($< 1\sigma$) |

**Table 4:** Results of the $\mathscr{L}$–test, testing for consistency of $BR(\overline{B} \to X_s\gamma)$ with all other data (excluding $\delta(g-2)_\mu$) and for consistency of $\delta(g-2)_\mu$ with all other data (excluding $BR(\overline{B} \to X_s\gamma)$).

both the flat and the CCR mSUGRA prior. The significance is reduced to $\sim 1.6\sigma$ under the log prior.

It is interesting how the predictive probabilities are almost independent on the choice of priors on the model's parameters, thus indicating a remarkable robustness in the model's

| Prior | $\ln \mathcal{L}(\delta(g-2)_\mu, b \to s\gamma | D)$ | Interpretation |
|---|---|---|
| $\delta(g-2)_\mu$ from $e^+e^-$ data | | |
| Flat | $-5.99 \pm 0.13$ | Incompatible at 95.0% significance |
| Log | $-5.87 \pm 0.13$ | Incompatible at 94.7% significance |
| CCR mSUGRA | $-6.42 \pm 0.14$ | Incompatible at 96.0% significance |
| $\delta(g-2)_\mu$ from $\tau$ decay data | | |
| Flat | $-1.59 \pm 0.07$ | Not significant ($1.26\sigma$) |
| Log | $-1.70 \pm 0.07$ | Not significant ($1.30\sigma$) |
| CCR mSUGRA | $-1.62 \pm 0.07$ | Not significant ($1.27\sigma$) |

**Table 5:** Results of the $\mathcal{L}$–test, jointly testing $b \to s\gamma$ and $\delta(g-2)_\mu$ for mutual compatibility and compatibility with all other observations, $D$. We find that the $\delta(g-2)_\mu$ observation from $e^+e^-$ data is incompatible with $b \to s\gamma$ at the $\sim 95\%$ level, almost independently of the choice of prior. On the other hand, no significant tension is detected for the $\delta(g-2)_\mu$ measurement from $\tau$ decay data.

predictions.
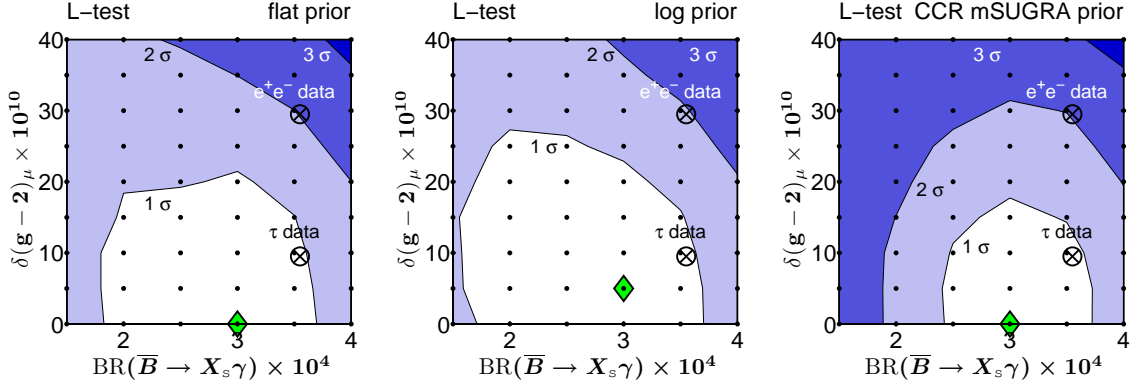


**Figure 2:** $\mathcal{L}$–test for both $b \to s\gamma$ and $\delta(g-2)_\mu$, for flat priors (left panel), log priors (middle panel) and CCR mSUGRA priors (right panel) in the CMSSM. The cross give the actual observed values (for the two different $\delta(g-2)_\mu$ determinations) and the green diamond is the most probable value under the model. Contours delimit values of $\ln \mathcal{L}(\delta(g-2)_\mu, b \to s\gamma | D) = 2.3, 6.17, 11.80$, corresponding to joint $1, 2, 3\sigma$ significance regions. The black, small dots indicate the locations at which the predictive probability has been evaluated, while the contours are interpolated.

We now consider the case where we test both $b \to s\gamma$ and $\delta(g-2)_\mu$ jointly, conditional on all other data. The result for the $\mathcal{L}$–test with $\mathcal{D} = (b \to s\gamma, \delta(g-2)_\mu)$ is shown in fig. 2 and reported in table 5. We can see that the CMSSM, given the observations $D$, tends to prefer small corrections to $\delta(g-2)_\mu$, although less so in the case of the log prior which gives more weight to the low-mass region where SUSY corrections tend to be larger. The joint observation of $b \to s\gamma$ and the determination of $\delta(g-2)_\mu$ based on $\tau$ decay data lies within the $1\sigma$ region, and hence no significant tension is detected between these two datasets. However, the $e^+e^-$ data based determination of $\delta(g-2)_\mu$ shows a $\sim 2\sigma$

| Prior | $\ln \mathcal{R}(\delta(g-2)_\mu, b \to s\gamma)$ | Interpretation |
|---|---|---|
| $\delta(g-2)_\mu$ from $e^+e^-$ data | | |
| Flat | $-0.62 \pm 0.20$ | Inconclusive evidence |
| Log | $-2.04 \pm 0.20$ | Weak evidence for incompatibility |
| CCR mSUGRA | $-0.69 \pm 0.25$ | Inconclusive evidence |
| $\delta(g-2)_\mu$ from $\tau$ data | | |
| Flat | $0.17 \pm 0.19$ | Inconclusive evidence |
| Log | $-0.64 \pm 0.17$ | Inconclusive evidence |
| CCR mSUGRA | $-0.58 \pm 0.23$ | Inconclusive evidence |

**Table 6:** Results of the $\mathcal{R}$–test, giving the relative odds (Bayes factor) between the two hypotheses that $\delta(g-2)_\mu$ and $b \to s\gamma$ are mutually compatible (corresponding to $\ln \mathcal{R} > 0$) or that they are in tension with each other and/or the rest of the data, $D$ (corresponding to $\ln \mathcal{R} < 0$). The statistical interpretation is in accordance with Jeffreys' scale, given in table 1.

significance for incompatibility (compare top part of table 5), which indicates an emerging tension between the two observations. The r.h.s. panel of fig. 2 gives the result for the CCR mSUGRA prior, which as it is shown in ref. [27] prefers the focus point region and large gaugino masses, penalizing large $\tan\beta$ values. This implies that, under this prior, regions of parameter space are favoured where the decoupling of SUSY contributions to $\delta(g-2)_\mu$ occurs and where negative contributions of the chargino-stop loop to $b \to s\gamma$ are suppressed. (Notice how the $3\sigma$ significance region in this case is much more extended). Despite this, the above results hold essentially unchanged even for this choice of prior.

### 4.2 Results for the $\mathcal{R}$–test

Turning now to the $\mathcal{R}$–test, we summarize the results in table 6 and plot the outcome in fig. 3. The $\mathcal{R}$–test for $b \to s\gamma$ and $\delta(g-2)_\mu$ returns an inconclusive result for all choices of priors, except for the case of log prior and $\delta(g-2)_\mu$ based on $e^+e^-$ data alone, which instead shows weak evidence for incompatibility. The reason for this result can easily be understood by considering fig. 3, which shows that for almost all possible observed values of $b \to s\gamma$ and $\delta(g-2)_\mu$ is undecided. Regions of large positive SUSY contributions to $\delta(g-2)_\mu$ and large negative corrections to $b \to s\gamma$ would be favoured (upper left corner of fig. 3), while regions of large positive corrections to $b \to s\gamma$ and large $\delta(g-2)_\mu$ values are disfavoured (upper right corner). The relative size of those region is somewhat prior dependent. This comes about in an analogous fashion as for the toy model presented in the Appendix: the $\mathcal{R}$–test deems observed values to be compatible if they tend to come from "compensating" regions of parameter space. However, by comparing fig. 3 with fig. 2, it is apparent that in this context the $\mathcal{L}$–test is the more stringent of the two, while the $\mathcal{R}$–test remains quite lenient, at least given the current experimental error on $\delta(g-2)_\mu, b \to s\gamma$.

## 5. Conclusions

We have subjected the question of the mutual compatibility of $b \to s\gamma$ and $\delta(g-2)_\mu$ to a detailed scrutiny, employing two different statistical tests that look for possible inconsis-
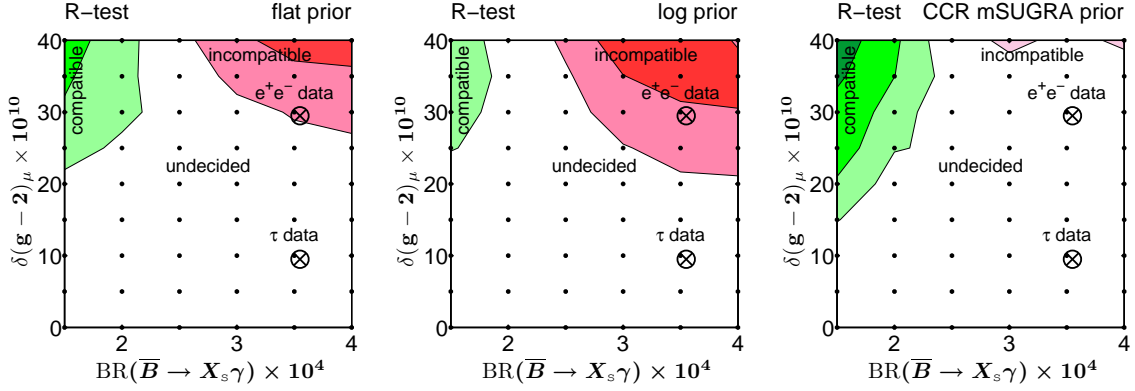
**Figure 3:** $\mathscr{R}$–test for both $BR(\overline{B} \to X_s\gamma)$ and $\delta(g-2)_\mu$, for the flat prior (left panel), log prior (middle panel) and CCR mSUGRA prior (right panel) in the CMSSM. The encircled crosses give the actual observed values (for the two different $\delta(g-2)_\mu$ determinations). Contours delimit values of $|\ln \mathscr{R}(\delta(g-2)_\mu, b \to s\gamma|D)| = 1.0, 2.5$, corresponding to levels of weak and moderate strength of evidence for either hypothesis, respectively, according to the Jeffreys' scale. The region in the top left corner favours $\mathcal{H}_0$ (that the two measurements are compatible), while the top right corner favours $\mathcal{H}_1$ (incompatible measurements). The white region returns an undecided result.

tencies between the two quantities and between the quantities and the model, in our case the CMSSM. We have found no sign of tension between $b \to s\gamma$ and the $\tau$ decay derived measurement of $\delta(g-2)_\mu$ under either test. On the other hand, our most stringent test shows a $\sim 95\%$ indication of tension between $b \to s\gamma$, the $e^+e^-$–based value of $\delta(g-2)_\mu$ and the other observed data (including the WMAP 5–yr dark matter determination). This can be interpreted in two ways: either as a sign of undetected systematics in the $e^+e^-$ value of $\delta(g-2)_\mu$, or (perhaps more interestingly) as an early indication of the difficulty of the CMSSM to simultaneously explain the observed values of $b \to s\gamma$ and $\delta(g-2)_\mu$. If the $\sim 3\sigma$ discrepancy in the anomalous magnetic moment of the muon is confirmed, this could be interpreted as evidence against the viability of the CMSSM.

## A. Illustration of the consistency tests on a toy problem

In order to illustrate the use of the Bayesian evidence to quantify the consistency between different datasets as discussed in section 2, we apply the $\mathscr{L}$–test (defined in eq. (2.5)) and the $\mathscr{R}$–test (defined in eq. (2.7)) to the simple linear problem of fitting a straight line through a set of data points, and then check for the consistency of a new observation with the previous measurements and with the model.

### A.1 Toy problem

We consider that the true underlying model for some process is a straight line described by

$$y(x) = mx + c, \tag{A.1}$$

where the free parameters in the model are the slope $m$ and the intercept $c$, whose true value is assumed to be 1 for both. The data consist of observations $y_i$ at known locations $x_i$, with Gaussian noise of known variance $\sigma$

$$y_i - y(x_i) = \epsilon \sim \mathcal{N}(0, \sigma). \tag{A.2}$$

We split the full dataset $d$ in two parts, $d = \{\mathscr{D}, D\}$, and we wish to test for the consistency of the subset $\mathscr{D}$ with the assumed subset $D$ and with the model of eq. (A.1). The likelihood function can then be written as

$$\mathcal{L}(m, c) \equiv p(d|m, c) = \prod_i p(d_i|m, c), \tag{A.3}$$

where

$$p(d_i|m, c) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-\chi_i^2/2] \tag{A.4}$$

and

$$\chi_i^2 = \sum_j \frac{(y(x_i) - y_i)^2}{\sigma^2}, \tag{A.5}$$

where $y(x_i)$ is the predicted value of $y$ at a given $x_i$ as a function of $c, m$ and $y_i$ is the measured value. We impose uniform, $\mathcal{U}(-5, 5)$ priors on both $m$ and $c$.

### A.2 Consistency tests for one observables at the time

In the first case, we wish to test for the consistency of one new observation with a set of previously gathered data points. We take the data set $D$ to consist of 9 data points at $x = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$ while the data set we wish to test, $\mathscr{D}$, consists of one observation at $x_9 = 9$. For definiteness, we use $\sigma = 0.5$ for the noise. We now employ the $\mathscr{L}$–test and the $\mathscr{R}$–test to check for consistency between datasets $D$ and $\mathscr{D} = y_9 \equiv y(x_9)$. We scan over the following $y_9$ values for $x = 9$

$$\mathscr{D}: \qquad y_9 = 7.5, \ldots, 12.5 \text{ in intervals of } 0.5. \tag{A.6}$$

The outcome of these two tests is shown in fig. 4, demonstrating that both tests correctly identify the region around $y_9 \sim 10$ as the one where the datasets are consistent
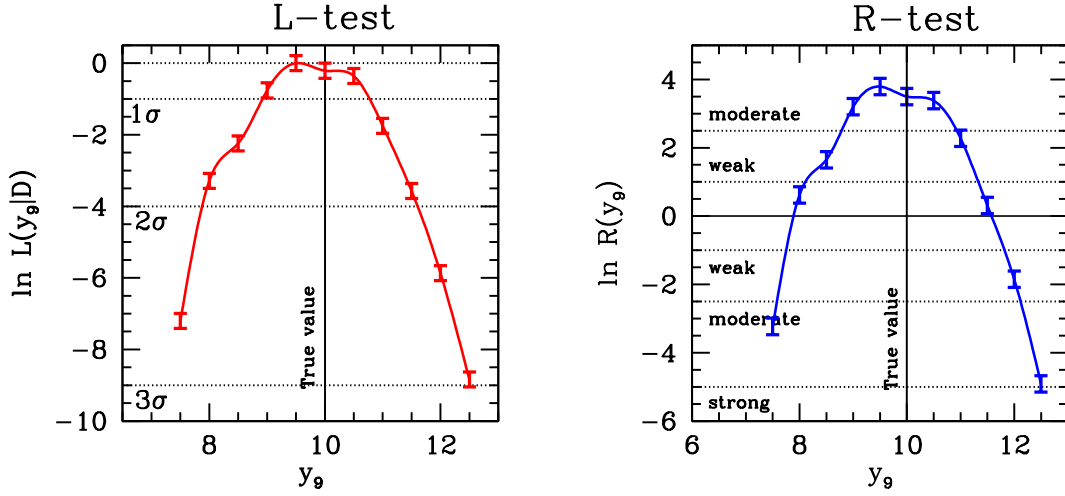
**Figure 4:** Left panel: toy model illustration of the 1–dimensional $\mathscr{L}$–test for $y_9$, with the horizontal, dashed lines representing levels of 1,2,3$\sigma$ significance. The vertical line is the value that corresponds to the true value of the model's parameters. Right panel: toy model illustration of the 1–dimensional $\mathscr{R}$–test. The horizontal lines delineate levels of evidence according to the Jeffreys' scale, in favour of compatibility (for $\ln \mathscr{R} > 0$) or against it (for $\ln \mathscr{R} < 0$). Both tests correctly identify the data region corresponding to the true model.

(notice that although the two curves look very similar they are not identical). According to the $\mathscr{R}$–test (right panel), the consistency hypothesis begins to be disfavoured in the regions $y_9 \lesssim 8.0$ and $y_9 \gtrsim 11.5$, which according to the $\mathscr{L}$–test corresponds to tension between the two datasets at the $\sim 2\sigma$ level (compare the left panel of fig. 4). As it is generically the case when comparing hypothesis testing using likelihood ratio and Bayesian model comparison, the significance levels of the former appear to give stronger results than the strength of evidence from the latter seems to justify. This is well known in the statistical literature, and in a particular version of this phenomenon goes under the name of "Lindley's paradox". For further details about interpreting and comparing the two results, see [26, 30] and references therein.

### A.3 Consistency tests for two observables jointly

In order to perform the consistency tests for two new observations jointly, we generate 8 data points at $x = \{0, 1, 2, 3, 4, 5, 6, 7\}$, again with Gaussian noise $\sigma = 0.5$. These data points are referred to as $D$. The dataset $\mathscr{D}$ now consists of $y(x = 8) \equiv y_8$ and $y(x = 9) \equiv y_9$. We scan over the following values for the possible outcomes of the observation for $\mathscr{D}$ (assuming the same noise properties as $D$):

$$y_8 = 6.5, \ldots, 11.5 \text{ in intervals of } 0.5 \tag{A.7}$$
$$y_9 = 7.5, \ldots, 12.5 \text{ in intervals of } 0.5.$$

The results of applying the $\mathscr{L}$–test and the $\mathscr{R}$–test are shown in fig. 5. It can be seen clearly from fig. 5 that the both tests favour the consistency hypothesis around the region with $y_8 \sim 9$ and $y_9 \sim 10$, which correspond to the outcome for the true value of the parameters (marked by an encircled black cross).
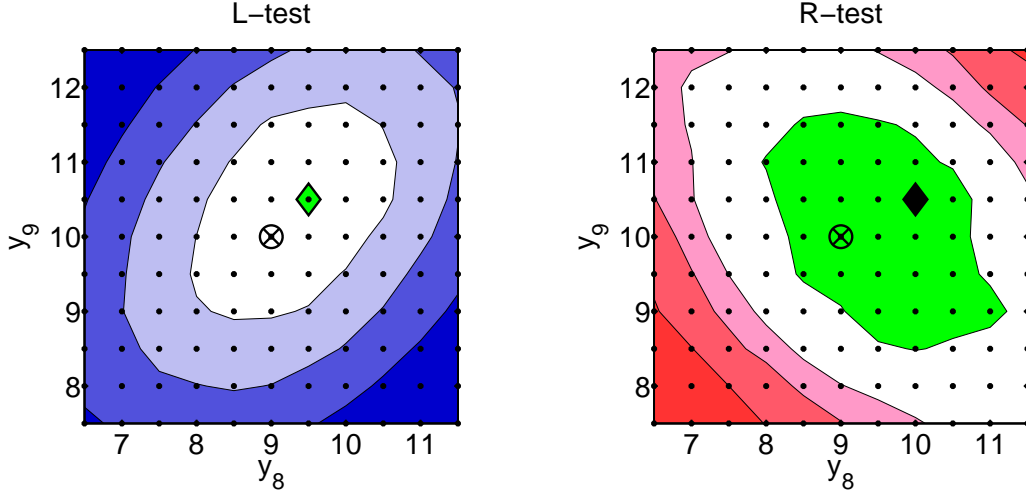


**Figure 5:** Left panel: toy model illustration of the 2–dimensional $\mathscr{L}$–test, showing the true value (black, encircled cross), the maximum probability prediction (green diamond) and 1,2,3$\sigma$ contours around it. Right panel: toy model illustration of the 2–dimensional $\mathscr{R}$–test. The black cross is the true value, while the black diamond is the point of maximum evidence in favour of the hypothesis of compatibility. Contours delineate regions of weak evidence in favour of compatibility (innermost/green region), inconclusive result (white), and regions of increasing evidence against compatibility (shades of red: weak, moderate and strong evidence according to the Jeffreys' scale).

It can also be seen from fig. 5 that although there is an overlap between different consistent and inconsistent regions favoured by the two tests, they generally prefer different regions as they look for inconsistency between datasets in a different manner as discussed in section 4. For this particular model of line fitting, the consistent region according to the $\mathscr{R}$–test is the one where the data points $y_8$ and $y_9$ lie on different sides of the true line given by eq. (A.1), i.e. the test favours either $y_8 > 9$ and $y_9 < 10$ or $y_8 < 9$ and $y_9 > 10$. The consistent region according to the $\mathscr{L}$–test is the one where both data points $y_8$ and $y_9$ lie on the same side of the true line, i.e. either $y_8 > 9$ and $y_9 > 10$ or $y_8 < 9$ and $y_9 < 10$. This difference between the two consistency tests can be understood by considering that the $\mathscr{R}$–test is trying to determine the probability that the datasets $\mathscr{D}$ and $D$ all come from the same model and so in order to enforce compatibility between them it favours the data points $y_8$ and $y_9$ to lie on different sides of the true line, thus preferring anti-correlated values. The $\mathscr{L}$–test, on the other hand, is trying to fit a straight line model for the given data values and so if $y_8$ is higher, it favours $y_9$ to be higher as well and vice versa, therefore favouring correlated behaviour.

# References

[1] See, e.g., S. P. Martin, *A Supersymmetry Primer*, arXiv:hep-ph/9709356.

[2] J. R. Ellis, S. Kelley and D. V. Nanopoulos, *Probing the desert using gauge coupling unification*, Phys. Lett. B **260** (1991) 131.

[3] E. Komatsu *et al.* [WMAP Collaboration], *Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations:Cosmological Interpretation*, Astrophys. J. Suppl. **180** (2009) 330 [arXiv:0803.0547 [astro-ph]].

[4] G. L. Kane, C. F. Kolda, L. Roszkowski and J. D. Wells, *Study of constrained minimal supersymmetry*, *Phys. Rev.* **D 49** (1994) 6173 [hep-ph/9312272].

[5] A. Chamseddine, R. Arnowitt and P. Nath, *Phys. Rev. Lett.* **49** (1982) 970;
R. Barbieri, S. Ferrara and C. Savoy, *Phys. Lett.* **B 119** (1982) 343;
L. J. Hall, J. Lykken and S. Weinberg, *Phys. Rev.* **D 27** (1983) 2359;
for a review, see, e.g., H. P. Nilles, *Supersymmetry, Supergravity and Particle Physics*, *Phys. Rept.* **110** (1984) 1;
A. Brignole, L. E. Ibañez and C. Muñoz, *Soft supersymmetry breaking terms from supergravity and superstring models*, published in *Perspectives on Supersymmetry*, ed. G. L. Kane, 125 [hep-ph/9707209].

[6] See, e.g., M. Drees and M. Nojiri, *The neutralino relic density in minimal N=1 supergravity*, *Phys. Rev.* **D 47** (1993) 376 [arXiv:hep-ph/9207234]; H. Baer and M. Brhlik, *Cosmological relic density from minimal supregravity with implications for collider physics*, *Phys. Rev.* **D 53** (1996) 597 [arXiv:hep-ph/9508321]; J. R. Ellis, T. Falk, G. Ganis, K.A. Olive and M. Srednicki, *The CMSSM parameter space at large tan beta*, *Phys. Lett.* **B 510** (2001) 236 [arXiv:hep-ph/0102098]; T. Nihei, L. Roszkowski and R. Ruiz de Austri, *New Cosmological and Experimental Constraints on the CMSSM*, *J. High Energy Phys.* **0108** (2001) 024 [arXiv:hep-ph/0106334]; A. Lahanas and V. Spanos, *Implications of the pseudo–scalar Higgs boson in determing the neutralino dark matter*, Euro. Phys. Journ. **23** (2002) 185 [arXiv:hep-ph/0106345].

[7] J. R. Ellis, S. Heinemeyer, K. A. Olive and G. Weiglein, JHEP **0301** (2003) 006 [arXiv:hep-ph/0211206].

[8] J. R. Ellis, S. Heinemeyer, K. A. Olive and G. Weiglein, *Phenomenological indications of the scale of supersymmetry*, *J. High Energy Phys.* **0605** (2006) 005 [hep-ph/0602220].

[9] J. Ellis, T. Hahn, S. Heinemeyer, K. A. Olive and G. Weiglein, JHEP **0710** (2007) 092 [arXiv:0709.0098 [hep-ph]].

[10] L. Roszkowski, R. Ruiz de Austri and R. Trotta, *Implications for the Constrained MSSM from a new prediction for $b \rightarrow s\gamma$*, *J. High Energy Phys.* **0707** (2007) 075 [arXiv:0705.2012].

[11] O. Buchmueller *et al.*, *Predictions for Supersymmetric Particle Masses in the CMSSM using Indirect Experimental and Cosmological Constraints*, arXiv:0808.4128 [hep-ph].

[12] E. A. Baltz and P. Gondolo, *Markov chain monte carlo exploration of minimal supergravity with implications for dark matter*, *J. High Energy Phys.* **0410** (2004) 052 [hep-ph/0407039].

[13] B. C. Allanach and C. G. Lester, *Multi-dimensional MSUGRA likelihood maps*, *Phys. Rev.* **D 73** (2006) 015013 [hep-ph/0507283].

[14] B. C. Allanach, *Naturalness priors and fits to the constrained minimal supersymmetric standard model, Phys. Lett.* **B 635** (2006) 123 [hep-ph/0601089].

[15] R. Ruiz de Austri, R. Trotta and L. Roszkowski, *A Markov Chain Monte Carlo analysis of the CMSSM, J. High Energy Phys.* **0605** (2006) 002 [hep-ph/0602028].

[16] B. C. Allanach, C. G. Lester and A. M. Weber, *The dark side of mSUGRA, J. High Energy Phys.* **0612** (2006) 065 [hep-ph/0609295].

[17] L. Roszkowski, R. Ruiz de Austri and R. Trotta, *On the detectability of the CMSSM light Higgs boson at the Tevatron, J. High Energy Phys.* **0704** (2007) 084 [hep-ph/0611173].

[18] R. Trotta, R. Ruiz de Austri and L. Roszkowski, *Prospects for direct dark matter detection in the constrained MSSM*, New Astron. Rev. **51** (2007) 316 [astro-ph/0609126].

[19] L. Roszkowski, R. Ruiz de Austri, J. Silk and R. Trotta, *On prospects for dark matter indirect detection in the Constrained MSSM*, arXiv:0707.0622 [hep-ph].

[20] B. C. Allanach, M. J. Dolan and A. M. Weber, *Global Fits of the Large Volume String Scenario to WMAP5 and Other Indirect Constraints Using Markov Chain Monte Carlo*, arXiv:0806.1184 [hep-ph].

[21] B. C. Allanach and D. Hooper, *Panglossian Prospects for Detecting Neutralino Dark Matter in Light of Natural Priors*, arXiv:0806.1923 [hep-ph].

[22] L. Roszkowski, R. R. de Austri, R. Trotta, Y. L. Tsai and T. A. Varley, *Some novel features of the Non-Universal Higgs Model*, arXiv:0903.1279 [hep-ph].

[23] F. Feroz and M. P. Hobson *Multimodal nested sampling: an efficient and robust alternative to MCMC methods for astronomical data analysis*, Mon. Not. Roy. Astron. Soc. **384** 449 (2008); F. Feroz, M. P. Hobson and M. Bridges, *MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics* (2008), arXiv:0809.3437.

[24] F. Feroz, B. C. Allanach, M. Hobson, S. S. AbdusSalam, R. Trotta and A. M. Weber, *Bayesian Selection of sign(mu) within mSUGRA in Global Fits Including WMAP5 Results, J. High Energy Phys.* **10** (2008) 064, arXiv:0807.4512.

[25] R. Trotta, F. Feroz, M.P. Hobson, L. Roszkowski and R. Ruiz de Austri, *The impact of priors and observables on parameter inferences in the Constrained MSSM, J. High Energy Phys.* **12** (2008) 024.

[26] R. Trotta, *Bayes in the sky: Bayesian inference and model selection in cosmology, Contemporary Physics*, **49**, 2, 71-104 (2008) [arXiv:0803.4089].

[27] M. E. Cabrera, J. A. Casas and R. R. de Austri, *Bayesian approach and Naturalness in MSSM analyses for the LHC*, arXiv:0812.0536 [hep-ph].

[28] K. L. Chan, U. Chattopadhyay and P. Nath, *Phys. Rev.* **D 58** (1998) 096004 [hep-ph/9710473].

[29] J. L. Feng, K. T. Matchev and T. Moroi, *Phys. Rev. Lett.* **84** (2000) 2322 [hep-ph/9908309] and *Phys. Rev.* **D 61** (2000) 075005 [hep-ph/9909334].

[30] R. Trotta, *Applications of Bayesian model selection to cosmological parameters*, Mon. Not. R. Astron. Soc. **378** (2007) 72 [astro-ph/0504022v3].

[31] G. D. Starkman, R. Trotta and P. M. Vaudrevange, *Introducing doubt in Bayesian model comparison* arXiv:0811.2415 [physics.data-an].

[32] P. Marshall, N. Rajguru and A. Slosar, *Bayesian evidence as tool for comparing datasets*, Phys. Rev. **D73** (2006) 067302; M. P. Hobson, S. L. Bridle and O. Lahav, *Combining cosmological data sets: hyperparameters and Bayesian evidence*, Mon. Not. Roy. Astron. Soc. 335, 377 (2002) [arXiv:astro-ph/0203259].

[33] C. Gordon and R. Trotta, *Bayesian Calibrated Significance Levels Applied to the Spectral Tilt and Hemispherical Asymmetry*, *Mon. Not. R. Astron. Soc.* **382** (2007) 1859-1863.

[34] J. P. Miller, E. de Rafael and B. L. Roberts, *Muon g-2: Review of Theory and Experiment*, *Rept. Prog. Phys.* **70** (2007) 795 [arXiv:hep-ph/0703049].

[35] M. Davier, S. Eidelman, A. Hocker and Z. Zhang, *Updated estimate of the muon magnetic moment using revised results from e+e- annihilation,* Eur. Phys. J. C **31** (2003) 503 [arXiv:hep-ph/0308213]. We use an updated value, as reported by M. Passera at a Workshop on the Topical Workshop on Muon Magnetic Dipole Moment $(g-2)_\mu$, 25 and 26 October 2007, Glasgow, UK, `http://www.ippp.dur.ac.uk/old/MuonMDM/`.

[36] Heavy Flavor Averaging Group (HFAG) (E. Barberio et al.), *Averages of b-hadron properties at the end of 2006*, arXiv:0704.3575 [hep-ex].

[37] M. Misiak and M. Steinhauser, *NNLO QCD corrections to the $\bar{B} \to X_s\gamma$ matrix elements using interpolation in $m_c$*, Nucl. Phys. **B 764** (2007) 62 [hep-ph/0609241].

[38] M. Misiak et al., *Estimate of $\bar{B} \to X_s\gamma$ at $\mathcal{O}(\alpha_s^2)$*, Phys. Rev. Lett. **98** (2007) 022002, [hep-ph/0609232].

[39] C. Degrassi, P. Gambino and G. F. Giudice, *$B \to X_s\gamma$ in supersymmetry: large contributions beyond the leading order*, *J. High Energy Phys.* **0012** (2000) 009 [hep-ph/0009337].

[40] P. Gambino and M. Misiak, *Quark mass effects in $\bar{B} \to X_s\gamma$*, Nucl. Phys. **B 611** (2001) 338 [hep-ph/0104034].

[41] K. Okumura and L. Roszkowski, *Deconstraining supersymmetry from $b \to s\gamma$*, Phys. Rev. Lett. **92** (2004) 161801 [hep-ph/0208101]; *Large beyond leading order effects in $b \to s\gamma$ in supersymmetry with general flavour mixing*, *J. High Energy Phys.* **0310** (2003) 024 [hep-ph/0308102].

[42] J. Foster, K. Okumura and L. Roszkowski, *New Higgs effects in B–physics in supersymmetry with general flavour mixing*, Phys. Lett. **B 609** (2005) 102 [hep-ph/0410323] and *Probing the flavour structure of supersymmetry breaking with rare B–processes: a beyond leading order analysis*, *J. High Energy Phys.* **0508** (2005) 094 [hep-ph/0506146].

[43] The CDF Collaboration, *Measurement of the $B_s - \bar{B}_s$ oscillation frequency*, Phys. Rev. Lett. **97** (2006) 062003 [hep-ex/0606027] and *Observation of $B_s - \bar{B}_s$ oscillations*, Phys. Rev. Lett. **97** (2006) 242003 [hep-ex/0609040].

[44] J. Dunkley et al. [The WMAP Collaboration], *Five-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Likelihoods and parameters from the WMAP data*, arXiv:0803.0586 [astro-ph].

[45] The CDF Collaboration, *Search for $B_s \to \mu^+\mu^-$ and $B_d \to \mu^+\mu^-$ decays in $p\bar{p}$ collisions with CDF-II*, CDF note 8956 (August 2007).

[46] The LEP Higgs Working Group, `http://lephiggs.web.cern.ch/LEPHIGGS`; G. Abbiendi et al. [the ALEPH Collaboration, the DELPHI Collaboration, the L3 Collaboration and the OPAL Collaboration, The LEP Working Group for Higgs Boson

Searches], *Search for the standard model Higgs boson at LEP*, *Phys. Lett.* **B 565** (2003) 61 [hep-ex/0306033].

[47] By CDF Collaboration and D0 Collaboration, *A Combination of CDF and D0 Results on the Mass of the Top Quark* arXiv:0803.1683 [hep-ex].

[48] W.-M. Yao et al., *The Review of Particle Physics*, *J. Phys.* **G33** (2006) 1 and 2007 partial update for the 2008 edition.

[49] K. Hagiwara, A. D. Martin, D. Nomura and T. Teubner, *Improved predictions for g-2 of the muon and* $\alpha_{\rm QED}(M_Z^2)$, *Phys. Lett.* **B 649** (2007) 173 [hep-ph/0611102].

[50] See `http://lepewwg.web.cern.ch/LEPEWWG`.